

INVESTIGATIONES LINGUISTICAE VOL. XXV
© INSTITUTE OF LINGUISTICS, ADAM MICKIEWICZ UNIVERSITY,
AL. NIEPODLEGŁOŚCI 4, 60-874, POZNAŃ, POLAND

Dystynktywność ciągów

Distinctiveness of sequences

Włodzimierz Lapis

INSTYTUT JĘZYKOZNAWSTWA, UNIWERSYTET IM. ADAMA MICKIEWICZA
AL. NIEPODLEGŁOŚCI 4, 61-874 POZNAŃ

lapisw@amu.edu.pl

Abstract

The article discusses the problem of sequence distinctiveness. The author sets out to define this concept and explain why the problem deserves attention. He then goes out to carry out a study investigating the degree of distinctiveness of one and two sequences (as a result of their comparison). The paper's framework is based on the properties of Pascal's triangle.

W artykule omawia się zagadnienie dystynktywności ciągów: wychodząc od zdefiniowania tego pojęcia i wykazania potrzeby zajęcia się tym zagadnieniem, przechodzi się do przeprowadzenia konkretnych badań, a mianowicie stopnia dystynktywności jednego jak i dwóch ciągów (w wyniku ich porównania). W pracy tej aktywnie wykorzystuje się własności trójkąta Pascala.

1. O potrzebie badania dystynktywności ciągów

Znakom przysługują następujące cechy¹:

1. znak jest dowolny (zasada arbitralna);
2. zmienność i niezmiennność znaku:
 - zmienny – z diachronicznego punktu widzenia (biorąc pod uwagę jego ewolucję przez wieki), może się zmienić forma, treść lub i to i to (bada się zmiany które zaszły od A do B),
 - niezmienny – z synchronicznego punktu widzenia (cecha charakterystyczna dla określonej chwili);
3. linearność (liniowość) – części znaku są usytuowane w określonej kolejności (kolejność ta jest cechą dystynktywną - odróżniającą spośród

¹ Źródło: <http://docs9.chomikuj.pl/82048979,0,1,J%C4%99zykoznawstwo-og%C3%B3lne,-NOTATKI-%28cz%C4%99C5%9B%C4%87-1%29.doc> (dostęp: 24 II 2011)

innych znaków), znaki tworzą układ na osi poziomej, jeden element następuje po drugim, np. k, o, t, = tok, kto, kot.

Ostatni z tych punktów wskazuje na istotną rolę kolejności znaków – ich kolejności w układzie linearnym, a więc że w pewnych sytuacjach nie tworzą one zbiorów, lecz ciągi.

Jest tak m.in. w przypadku w przypadku języka. Językowy przekaz informacji ma bowiem formę ciągu znaków mowy czy tekstu. Jeśli więc chcielibyśmy do niego odnieść pojęcie dystynktywności, to wówczas musielibyśmy rozpatrzeć pojęcie dystynktywności w odniesieniu nie do elementów czy zbioru elementów, lecz właśnie do ciągu elementów.

Zastanówmy się kiedy możemy mieć do czynienia z tego typu sytuacjami. W tym celu zauważmy, że gdy chociażby porównujemy dwa teksty (choćby w celu **wykrycia plagiatu** lub sprawdzając **w jakim stopniu fraza z wyszukiwarki internetowej pokrywa się ze znalezionym przez nią wynikiem**) – sprawdzamy stopień ich bliskości (podobieństwa) lub różności, oddalenia (tj. właśnie dystynktywności). Analogicznie możemy też rozpatrywać chociażby kwestie **porównywania DNA** (jako ciągu nukleotydów) czy **rozpoznawania tekstu za pomocą programu OCR** (a więc porównywania kolejnych znajdujących się w nim liter ze znajdującymi się w nim ich wzorcami). Wszst

Ciąg (zgodnie z definicją tego pojęcia) rozumiemy jako funkcję:

- $f: N \rightarrow Y$ gdy jest to ciąg nieskończony,
- $f: \{1, 2, 3, \dots, n\} \rightarrow Y$ gdy jest to ciąg skończony (o liczbie elementów n),

gdzie Y jest dowolnym zbiorem elementów.

W naszym zaś przypadku, gdy rozpatrujemy jedynie obiekty binarne, będzie to funkcja:

- $f: N \rightarrow \{0,1\}$ gdy jest to ciąg nieskończony,
- $f: \{1, 2, 3, \dots, n\} \rightarrow \{0,1\}$ gdy jest to ciąg skończony (o liczbie elementów n).

Ciągi takie nazywać będziemy ciągami binarnymi, i dalej tylko o takich ciągach będziemy tu mówić (podobnie jak w pozostałych częściach tej pracy ograniczymy się do obiektów binarnych).

Właściwie ograniczymy się tylko do ciągów skończonych, jako że każda z wypowiedzi też jest skończona.

1.1 Klasyczna definicja stopnia dystynktywności dwóch ciągów

Jeśli dane nam będzie rozpatrywać dwa ciągi, każdy o długości n , to wówczas stopień ich dystynktywności – to stopień ich różności, określający na ilu z tych n pozycji się one różnią.

01101	}	te 2 ciągi różnią się na dwóch pozycjach, więc stopień ich dystynktywności wynosi 2.
01110		

Gdy weźmiemy – damy na to – ciąg 3-elementowy i będziemy go porównywać z innym ciągiem 3-elementowym, to takich sytuacji będziemy mieć $\binom{8}{2} = 28$. Ciągów 3-elementowych jest bowiem 8, a z nich wybieramy 2, przy czym mają to być różne ciągi i nie ważna jest ich kolejność – stąd właśnie kombinacja $\binom{8}{2}$ (w ogólnej sytuacji będzie ich oczywiście $\binom{2^n}{2}$, gdzie n określa długość ciągu).

Wypisując wszystkie te sytuacje, otrzymujemy że mamy:

- 4 sytuacje 0N 3R (tj. takie w których na 0 pozycji się Nie różnią, a na 3 się Różnią),
- 12 sytuacji 1N 2R,
- 12 sytuacji 2N 1R

(stosunek liczby tych sytuacji jest więc następujący: 1:3:3).

Gdy z kolei weźmiemy wszystkie możliwe zestawienia ciągów 3-elementowych (tzw. wariacje z powtórzeniami) – to będzie ich 8^3 – będą to powyższe sytuacje razy dwa, bo istotne będzie który ciąg z którym porównujemy (pierwszy z drugim czy drugi z pierwszym) oraz porównania danego ciągu z nim samym (a w ogólnej sytuacji będzie ich więc $2 \cdot 8^3$).

Dla omawianej tu sytuacji porównywania dwóch ciągów 3-elementowych, otrzymujemy następujący układ zestawień tych porównań:

- 8 sytuacji 0N 3R,
- 24 sytuacje 1N 2R,
- 24 sytuacje 2N 1R,
- 8 sytuacji 3N 0R.

Ich stosunek jest więc 1:3:3:1, a więc utworzony jest z liczb, które kolejno znajdziemy w 3. wierszu w trójkącie Pascala.

Trójkąt Pascala – to trójkątna tablica liczb:

0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	</
---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

.....

utworzona w ten sposób, że na bokach owego trójkąta znajdują się liczby 1, a każda z pozostałych znajdujących się w nim liczb powstaje jako suma dwóch bezpośrednio znajdujących się nad nią liczb. Jeżeli przyjmiemy, że zarówno kolumny, wiersze, jak i znajdujące się w nich elementy, będziemy liczyć od zera, to wówczas możemy powiedzieć, że liczby stojące w n -tym wierszu, to kolejne wartości dwumianu Newtona $\binom{n}{k}$ o owym n i o k równym kolejno od 0 do n (co oznacza że liczba stojąca na

miejscu k w wierszu n jest równa $\binom{n}{k}$), będące zarazem współczynnikami kolejnych wyrazów rozwinięcia $(a + b)^n$, tj.:

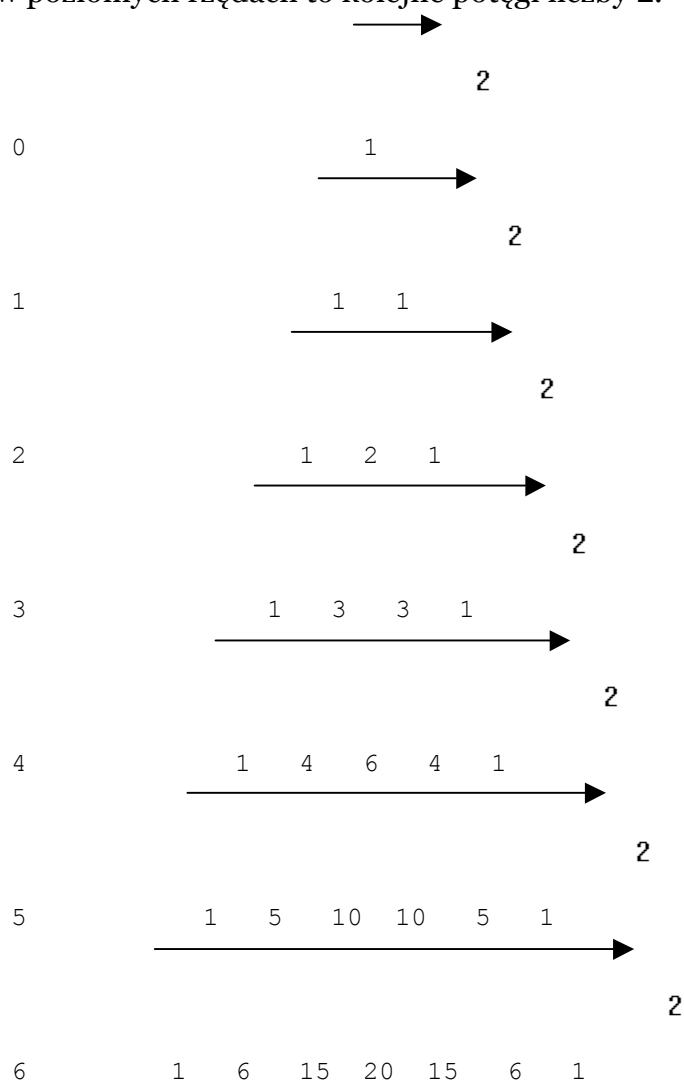
$$(a+b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b^1 + \dots + \binom{n}{k} a^{n-k} b^k + \dots + \binom{n}{n-1} a^1 b^{n-1} + \binom{n}{n} a^0 b^n$$

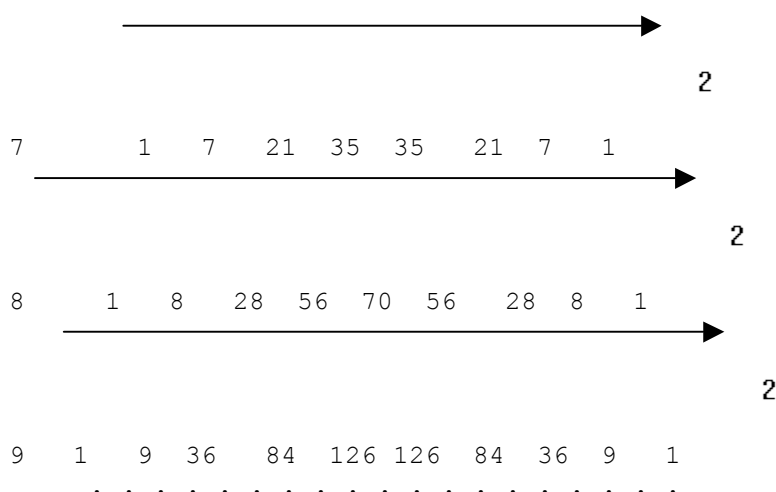
Stąd mamy:

$$(a+b)^3 = \binom{3}{0}a^3b^0 + \binom{3}{1}a^2b^1 + \binom{3}{2}a^1b^2 + \binom{3}{3}a^0b^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

co oznacza, że w trzecim wierszu trójkąta Pascala mamy właśnie kolejno liczby: 1, 3, 3, 1.

Tego typu trójkąt ma bardzo wiele interesujących własności. Dla nas jednak, oprócz wyżej wymienionych, w prowadzonych tu rozważaniach, jako ważna i interesująca, jest jedynie własność, która mówi że: sumy liczb w poziomych rzędach to kolejne potęgi liczby 2.





$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

Ilustruje to powyższy rysunek oraz następujący wzór:

Na boku zauważmy tu, że ciągi **010101** i **101010** są bardzo do siebie podobne (choć różnią się na każdej z pozycji). Po ich przesunięciu względem siebie, otrzymalibyśmy bowiem, że ich stopień podobieństwa wynosi 5: **010101** i **101010**. Analogiczną sytuację możemy zresztą rozpatrywać mając większą liczbę elementów. Sprawdzamy wówczas maksymalnie na ilu pozycjach się one

wspólnie pokrywają przy pewnym ich przesunięciu:

(w tym przypadku oczywiście na 3 pozycjach).

Nie interesuje więc nas choćby następująca sytuacja:

101111
111001
0001, gdyż jesteśmy zainteresowani, by wszystkie te ciągi pokrywały się wspólnie na tym samym ciągu wartości.

Przy tym, fragmenty na których te ciągi pokrywają się, wcale nie muszą być spójne, gdyż mogą przecież pokrywać się na kilku składowych

spójności: **11100010010**
01101110100
110011000010 (tu akurat mamy 3 składowe spójności).

W dalszych jednak rozważaniach, omawiając zagadnienie dystynktywności ciągów, będziemy korzystać z podanej wcześniej klasycznej definicji stopnia dystynktywności dwóch ciągów.

2. Badanie rozkładu stopnia dystynktywności dwóch ciągów

Wróćmy jednak do głównego nurtu naszych rozważań. Na początku, rozpatrując wszystkie możliwe przypadki, sprawdzamy jeśli będziemy porównywać dwa ciągi o równej długości, to na ilu pozycjach mogą się odpowiednio: różnić czy zgadzać. Dla przykładu sprawdzimy to dla ciągów 3-elementowych, jednak na każdym kroku od razu będziemy dokonywać uogólnienia dla niepustych ciągów o dowolnej skończonej długości. W rozważaniach tych, ze względu na tematykę tej pracy, wyjdziemy od kwestii różnienia się (a więc dystynktywności) ciągów, a następnie, mając już to zrobione, będzie już nam bardzo łatwo uporać się z drugim z tych zagadnień, jakim jest kwestia zgadzania się porównywanych ciągów.

W tym celu przypatrzmy się poniższej tabeli, w której zestawiono porównanie między sobą wszystkich binarnych ciągów 3-elementowych:

Te ciągi (bazowe)	Porównujemy z tymi ciągami (na szarym tle) i zapisujemy (na jasnym tle) jak się z nimi różnią na poszczególnych pozycjach (1 = różnią się, 0 = nie różnią się)							
	000	001	010	011	100	101	110	111
000	000	001	010	011	100	101	110	111
001	001	000	011	010	101	100	111	110
010	010	011	000	001	110	111	100	101
011	011	010	001	000	111	110	101	100
100	100	101	110	111	000	001	010	011
101	101	100	111	110	001	000	011	010
110	110	111	100	101	010	100	000	001
111	111	110	101	100	011	010	001	000

Tab. 1

Po dokładniejszym przestudiowaniu otrzymanych wyników, widzimy że wśród nich, zarówno w każdym wierszu, jak i w każdej kolumnie, znajdują się wszystkie możliwe 3-elementowe kombinacje zer i jedynek (każda po 1 raz).

Jeśli teraz dodatkowo zliczymy jedynki w poszczególnych ciągach wynikowych tabeli 1, to otrzymamy poniższą tabelę 2a, informującą nas o stopniu dystynktywności porównywanych ciągów, oraz otrzymaną z niej tabelę 2b, w której znajdują się zliczenia tych sytuacji (wierszami i u dołu całościowo).

Te ciągi (bazowe)	Porównujemy z tymi ciągami (na szarym tle) i zapisujemy (na jasnym tle) na ilu pozycjach się różnią:								Na białym tle zliczamy na ilu pozycjach się różnią:			
	000	001	010	011	100	101	110	111	0	1	2	3
000	0	1	1	2	1	2	2	3	1	3	3	1
001	1	0	2	1	2	1	3	2	1	3	3	1
010	1	2	0	1	2	3	1	2	1	3	3	1
011	2	1	1	0	3	2	2	1	1	3	3	1
100	1	2	2	3	0	1	1	2	1	3	3	1
101	2	1	3	2	1	0	2	1	1	3	3	1
110	2	3	1	2	1	2	0	1	1	3	3	1
111	3	2	2	1	2	1	1	0	1	3	3	1
RAZEM →									8	24	24	8

Tab. 2a

Tab. 2b

W oparciu o tab. 2b widzimy, że w tab. 2a:

- w każdej jej linii jest taki sam (liczebnie) rozkład różniących się wystąpień na danej pozycji,
- rozkład ten ma postać: 1-3-3-1.

Poniżej pokażemy dlaczego zachodzą powyższe dwie własności. W tym celu:

- 1) najpierw pokażemy dlaczego taki właśnie rozkład jest w pierwszej linii powyższej tabeli (gdyż właśnie w odniesieniu do niej najłatwiej jest to zrobić),
- 2) a następnie pokażemy dlaczego identyczny z nim jest rozkład w każdym z pozostałych wierszy tejże tabeli.

Ad 1)

W tym celu musimy się odwołać do tab. 1. W jej pierwszej linii porównujemy ciąg 000 z wszelkimi możliwymi różnymi ciągami o 3 pozycjach. Widzimy, że ciąg 000 każdorazowo różni się z poszczególnymi ciągami z nagłówek kolejnych kolumn tejże tabeli jedynie jedynekami. Jedyńki zaś (co do krotności liczby ich wystąpień w poszczególnych ciągach w nagłówkach kolumn) rozkładają się zgodnie z wartościami kolejnych dwumianów Newtona:

- $\binom{3}{0}$ - to liczba 0 wystąpień jedyńki na 3 pozycjach, wynosi ona 1,
- $\binom{3}{1}$ - to liczba 1 wystąpienia jedyńki na 3 pozycjach, wynosi ona 3,
- $\binom{3}{2}$ - to liczba 2 wystąpień jedyńki na 3 pozycjach, wynosi ona 3,
- $\binom{3}{3}$ - to liczba 3 wystąpień jedyńki na 3 pozycjach, wynosi ona 1,

a co za tym idzie (ze względu na jedną z własności trójkąta Pascala) – są to kolejne liczby z trzeciego jego wiersza:

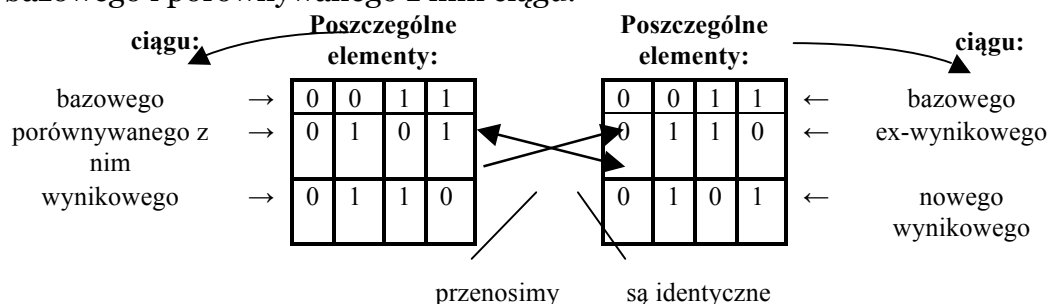
1
1 1
1 2 1
1 3 3 1

Oczywiście z analogiczną sytuacją będziemy mieli do czynienia dla ciągów o innej skończonej liczbie elementów. □

Ad 2)

Aby pokazać, że dla ciągów 3-elementowych, identyczny rozkład różnienia się na danej liczbie pozycji (a mianowicie: 1-3-3-1) otrzymujemy również w każdej z pozostałych linii, tj. w wyniku porównywania innego „ciągu bazowego” z wszelkimi różnymi możliwymi ciągami o tej samej długości, wystarczy pokazać, że w tab. 1, analogiczne jak w wierszu 1, również w pozostałych wierszach, po jednym razie występują wszystkie możliwe binarne ciągi o długości 3 (dla ciągów o dowolnej skończonej długości n robi się to analogicznie).

Aby jednak to zrobić, przyjrzyjmy się poniższym zasadom przekształcania 4 różnych układów pojedynczych elementów: ciągu bazowego i porównywanego z nim ciągu.



Schemat ten wymaga odpowiedniego opisu.

Po lewej stronie mamy wyszczególnione 4 możliwe sytuacje ustawień wartości elementu z ciągu bazowego (0 lub 1) i wartości elementu porównywanego z nim ciągu (też 0 lub 1), jak również odpowiadające im wartości elementu ciągu wynikowego (określone według zasad podanych w nagłówku tab. 1).

Następnie – po prawej stronie – mamy analogiczną sytuację, jednak tu porównujemy elementy tego samego ciągu bazowego z elementami przeniesionego ciągu wynikowego. Co ciekawe, elementy w ten sposób otrzymanego nowego ciągu wynikowego są identyczne jak elementy ciągu porównywanego z ciągiem bazowym po lewej stronie schematu.

Ciąg, to nic innego, jak zestawienie jego elementów, a zatem operowanie na nim, to zestawienie operacji na kolejnych jego elementach. Możemy więc powiedzieć, że ciąg bazowy spełnia tu funkcję swego rodzaju klucza kodującego-dekodującego:

- gdy go zestawimy z ciągiem porównywanym z nim – to (jak to ma miejsce po lewej stronie powyższego schematu) otrzymujemy ciąg wynikowy,
- a gdy go zestawimy z owym ciągiem wynikowym – to (jak to ma miejsce po prawej stronie powyższego schematu) otrzymujemy ten sam ciąg z którym przed chwilą porównywaliśmy ciąg bazowy (po lewej stronie schematu).

Ciąg porównywany i ciągi wynikowy, przy owym porównywaniu, są więc wzajemnie wymienialne za sprawą danego ciągu bazowego.

To bezpośrednio prowadzi nas do wniosku, że rzeczywiście w każdej linii tab. 1 znajdują się ciągi każdego rodzaju o danej długości (co pokazujemy poniżej).

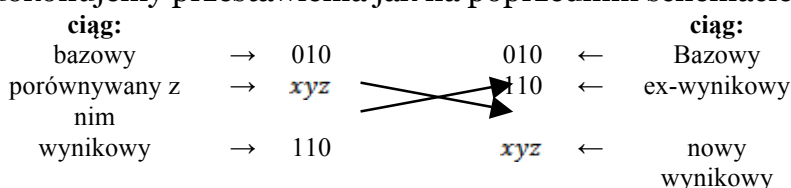
W tabeli tej za każdą linię odpowiada opisujący ją ciąg bazowy. Weźmy dla przykładu trzecią linię (dla pozostałych robi się analogicznie), a więc o ciągu bazowym 010. Chcemy teoretycznie sprawdzić („ręcznie” można bowiem bezpośrednio w owej tabeli), czy istnieje w tej linii dowolny ciąg binarny danej długości (np. 110), czyli czy da się go otrzymać przez porównanie ciągu bazowego z tej linii z JAKIMŚ 3-elementowym ciągiem

010

xyz

binarnym: 110 .

Aby się dowiedzieć jaki to powinien być ten „JAKIŚ” ciąg – dokonujemy przestawienia jak na poprzednim schemacie:



I poprzez porównanie po prawej stronie tego schematu – otrzymujemy, że jest to ciąg **100**. Tym samym w tab. 1 znaleźliśmy kolumnę, w której w badanym wierszu znajduje się ciąg 110. □

Dodajmy tu, że jako że nie jest istotne, czy porównujemy ciąg bazowy z porównywanym z nim, czy też porównujemy je w odwrotnej kolejności – zatem otrzymawszy właśnie że w każdym wierszu rozważanej tabeli znajdują się wszystkie możliwe ciągi o danej długości – wynik taki automatycznie otrzymujemy dla każdej z kolumn tejże tabeli.

Dodatkowo zauważmy, że skoro (jak dopiero co udowodniliśmy) w każdym wierszu tab. 2a jest rozkład dystynkcji jak to podaje tab. 2b, a więc identyczny, więc zliczając je – również otrzymujemy identyczny rozkład, a mianowicie:

- w przypadku rozważanych tam binarnych ciągów 3-elementowych: 1-3-3-1, a więc jak w 3. linii trójkąta Pascala, czy też jakie wartości mają kolejne wyrazy dwumianu Newtona $\binom{3}{k}$, gdzie k równa się kolejno: 0, 1, 2 i 3;
- w ogólnym przypadku ciągów binarnych n -elementowych: jak w n -tej linii trójkąta Pascala, czy też jakie wartości mają kolejne wyrazy dwumianu Newtona $\binom{n}{k}$, gdzie k równa się kolejno: 0, 1, 2, ..., n .

Obecnie – zgodnie z wcześniejszą zapowiedzią – możemy już pokrótce omówić zagadnienie stopnia zgadniania się porównywanych ciągów. Jak się jednak okazuje – co do rozkładu poszczególnych sytuacji – jest ono identyczne z zagadnieniem dystynktywności. Jest tak dlatego, że zawsze jeśli dane dwa ciągi na k pozycjach się różnią, to na $n - k$ pozycjach się nie różnią (i na odwrót), a rozkłady różnienia się na danej i dopełnieniowej liczbie pozycji są symetryczne (patrz tab. 2b).

3. Określanie / badanie stopnia dystynktywności ciągów

Weźmy jeden ciąg o n pozycjach i drugi, też o n pozycjach, jednak różniący się z tym pierwszym na k pozycjach. Poniżej nasze rozważania będziemy prowadzić równolegle na dwóch poziomach – na konkretnym przykładzie dla $n=8$ i $k=3$ oraz na sytuacji ogólnej. Oczywiście widzimy, że wystarczy wybrać te k ($=3$) pozycje spośród n ($=8$) pozycji na których mają się te dwa ciągi różnić. Możemy to zrobić na: w omawianej tu sytuacji szczególnej $\binom{8}{3} = 56$ sposobów, a w sytuacji ogólnej – na $\binom{n}{k}$ sposobów.

Poniżej sprawdzimy jak ten wynik liczbowo rozkłada się na sytuację gdy ustalony / wiadomy jest numer pierwszej pozycji ciągu na której się one różnią.

Jeśli będzie on na miejscu	To pozostałe 2 będą na pozostałych ... pozycjach:	Sytuacji takich jest więc:	W sytuacji ogólnej będzie ich oczywiście:
1.	7	$\binom{7}{2} = 21$	$\binom{n-1}{k-1}$
2.	6	$\binom{6}{2} = 15$	$\binom{n-2}{k-1}$
3.	5	$\binom{5}{2} = 10$	$\binom{n-3}{k-1}$
4.	4	$\binom{4}{2} = 6$	$\binom{n-4}{k-1}$
5.	3	$\binom{3}{2} = 3$	$\binom{n-5}{k-1}$
6.	2	$\binom{2}{2} = 1$	$\binom{k-1}{k-1}$ (tu $k-1 = n-6$)
SUMA WYNIKÓW:		$21 + 15 + 10 + 6 + 3 + 1 = 56$ czy też: $\sum_{k=2}^7 \binom{k}{2} = \binom{7+1}{2+1} = \binom{8}{3}$	$\sum_{l=k-1}^{n-1} \binom{l}{k-1} = \binom{n-1+1}{k-1+1} = \binom{n}{k}$

Do tego, że taka jest w rzeczywistości wartość owej sumy, można bardzo łatwo dojść korzystając z przytoczonych wcześniej własności trójkąta Pascala:

0
 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100

Mamy pokazać, że
elementy zaciennione
sumują się do
podkreślnego

W pierwszym kroku zastępujemy jedną jedynekę inną jedyneką (co nie wpływa na wynik)

[illegible]

W kolejnym kroku w miejsce dwóch górnych zacieniowanych elementów bierzemy znajdujący się bezpośrednio pod nimi element, a (zgodnie z własnościami trójkąta Pascala) będący ich sumą, a zatem i wartość całej sumy się nie zmienia

[illegible]

Powtarzamy ten krok

[illegible]

Powtarzamy ten krok

[illegible]

Powtarzamy ten krok

[illegible]

Powtarzamy ten krok

$$\begin{array}{ccccc} & & & & 1 \\ 0 & & & & \\ & & & & \\ 1 & & & 1 & 1 \end{array}$$

2					1	2	1							
3					1	3	3	1						
4					1	4	6	4	1					
5					1	5	10	10	5	1				
6					1	6	15	20	15	6	1			
7					1	7	21	35	35	21	7	1		
8					1	8	28	56	70	56	28	8	1	
9					1	9	36	84	126	126	84	36	9	1
				

W tym momencie
otrzymaliśmy już
szukany wynik.

Oczywiście proces ten można stosować dla dowolnych wartości n i k , co znaczy że ma on i zastosowanie dla „przypadku ogólnego”.

Przejdźmy obecnie do sprawdzenia, jeśli będziemy czytać nasze słowo (ciąg) od strony lewej do prawej, to **jaka jest wartość oczekiwana liczby sprawdzeń aż trafimy na literę (element) różnicującą(-cy).**

Tu nasze rozważania szczegółowe też będziemy prowadzić dla $n=8$ i $k=3$. W tym celu wykorzystamy poczynione już powyżej rozważania co do ilości sytuacji gdy pierwszy element jest już na określonym / znanym nam miejscu.

Podobnie jak poprzednio, sposobów rozmieszczenia owych 3 różnicujących elementów na 8 pozycjach jest oczywiście $\binom{8}{3} = 56$, a w sytuacji ogólnej będzie to oczywiście $\binom{n}{k}$.

Gdy element pierwszy jest na pozycji:	To sytuacji takich jest:	A zatem idąc od lewej strony, by trafić na ów pierwszy element – łącznie (we wszystkich tych sytuacjach) trzeba wykonać sprawdzeń:	To samo ogólnie:
1	$\binom{7}{2} = 21$	$1 \cdot 21 = 21$	$1 \cdot \binom{n-1}{k-1}$
2	$\binom{6}{2} = 15$	$2 \cdot 15 = 30$	$2 \cdot \binom{n-2}{k-1}$
3	$\binom{5}{2} = 10$	$3 \cdot 10 = 30$	$3 \cdot \binom{n-3}{k-1}$
4	$\binom{4}{2} = 6$	$4 \cdot 6 = 24$	$4 \cdot \binom{n-4}{k-1}$
5	$\binom{3}{2} = 3$	$5 \cdot 3 = 15$	$5 \cdot \binom{n-5}{k-1}$
6	$\binom{2}{2} = 1$	$6 \cdot 1 = 6$	$(n-k+1) \cdot \binom{k-1}{k-1}$
RAZEM:		126	$\sum_{i=k-1}^{n-i} \left((n-i) \cdot \binom{i}{k-1} \right)$
WARTOŚĆ OCZEKIWANA		$\frac{126}{56} = \frac{9}{4}$	$\frac{\sum_{i=k-1}^{n-i} \left((n-i) \cdot \binom{i}{k-1} \right)}{\binom{n}{k}}$

Zauważmy, że w sumie do dodania mamy tutaj następujące wartości:

21						}	trójkąt elementów wyjściowych
15	15						
10	10	10					
6	6	6	6				
3	3	3	3	3			
1	1	1	1	1	1		
S_1	S_2	S_3	S_4	S_5	S_6		

Jak już to stwierdziliśmy wyżej, pierwsza (od lewej) z tych kolumn sumuje się do 56, a zatem (w analogiczny sposób na trójkącie Pascala) i pozostałe kolumny sumują się odpowiednio do: 35, 20, 10, 4 i 1, co przedstawia się następująco:

[illegible]

Kolejne podsumy jasno zacięnięwanych elementůw dają nam w wyniku wartości podkreślone.

Gdy teraz zsumujemy owe podkreślone elementy – to otrzymamy wynik 126 (intensywniej zaciemniony element) będący sumą wcześniej obliczonych sum S_1 do S_6 , a więc (innymi słowy) – sumą wszystkich elementów z „trójkąta elementów wyjściowych”.

Tak więc jak przy wyborze 3 elementów z 8 operowaliśmy wartościami $\binom{8}{3} = 56$ i $\binom{9}{4} = 126$ – tak więc (ogólnie) przy wyborze k elementów z n , jako wyniki te otrzymujemy odpowiednio $\binom{n}{k}$ i $\binom{n+1}{k+1}$.

Gdy więc teraz będziemy chcieli obliczyć wartość oczekiwaną liczby ruchów do trafienia na pierwszy element różnicujący, to otrzymamy:

$$\frac{126}{56} = \frac{9}{4}$$

1) w sytuacji $n=8$ i $k=3$: ;

2) w sytuacji ogólnej:

$$\frac{\binom{n+1}{k+1}}{\binom{n}{k}} = \frac{\frac{(n+1)!}{(k+1)! \cdot (n+1-k-1)!}}{\frac{n!}{k! \cdot (n-k)!}} = \frac{\frac{n! \cdot (n+1)}{k! \cdot (k+1) \cdot (n-k)!}}{\frac{n!}{k! \cdot (n-k)!}} = \frac{n+1}{k+1}$$

(i) rzeczywiście, gdy $n=8$ i $k=3$, to: $\frac{n+1}{k+1} = \frac{8+1}{3+1} = \frac{9}{4}$.

[illegible]

O co tu chodzi?!

Gdy porównujemy dwa obiekty (choćby w celu sprawdzenia czy są identyczne, czy choćby podobne) – wówczas sprawdzamy stopień ich bliskości (podobieństwa) lub różności, oddalenia (tj. właśnie dystynktywności). W szczególności możemy też rozpatrywać w tym względzie chociażby kwestie: rozpoznawania ludzi po odciskach palców (porównywani ich z odciskami z bazy danych), czy porównywania obrazów.

Przy rozpoznawaniu obiektu, w zależności od tego co robimy (jaką rzeczywistość realną badamy), istotne jest:

- czy obiekt ten znajduje się w znanej nam rzeczywistości (tj. znane są nam wszystkie jej obiekty i własności pozwalające nam odróżnić ów obiekt od pozostałych) – wtedy całość pracy może zachodzić jedynie w umyśle badacza,
- czy też cała rzecz dzieje się w dotychczas nie znanej nam (tj. „zastanej”) rzeczywistości – wówczas postępujemy jak powyżej, wcześniej jednak będziemy musieli określić własności tych obiektów, i to w takim stopniu, by za ich pomocą ów obiekt mógł już być wyróżniony spośród pozostałych obiektów tegoż uniwersum.

Jest to dystynktywność, (tj. „różność”) kontekstowa elementu względem zbioru.